Chrome Extension for Code Summarization

Team Member

Shan Huang - sh69@illinois.edu

Topic Choice

The topic **code summarization** is a task that generates descriptions directly from the source code. Good code summarization can be viewed as a form of document expansion, which could help with the performance of code searching that utilizes natural language queries and maintenance of the code. However, writing those code summaries manually is time-consuming, and therefore makes it critical and valuable to genearate code summaries automatically. However, this task requires a solid understanding of the code segments, which is difficult due to several reasons. For example, code structures are complex, involving various sequential and parallel structures. Besides, it is hard to set standards for a "good" code summary. In addition, the code styles are different for different people.

The theme that this problem belongs to is **Intelligent Browsing**, which is going to be delivered as a Chrome extension. Code summarization can be viewed as a text generation task with given source code and comments. This problem belongs to text mining, which is taught in the class. In addition, language models learned in class, which compute the probability of each word, are used frequently in this kind of problem.

Datasets and Algorithms

Use model on this github repo. Model trained on dataset

Demonstration

Demo the extension by showing how it behaves to some code segmentations from the dataset.

Found limitations of the model through testing.

Programming Language

Mainly Python (Pytorch)

Workload Justification

1 person time -> at least 20 hours

Task	Time Expected
Understanding and run the Model	5
Implementing chrome extension with python model	15
Testing the extension and found limitation of model	5